Quantitative neuroanatomy for connectomics in *Drosophila*

Casey M. Schneider-Mizell[1*], Stephan Gerhard[1,2*], Mark Longair[2],
Tom Kazimiers[1], Feng Li[1], Maarten F. Zwart[1], Andrew Champion[1],
Frank Midgley[1], Richard Fetter[1], Stephan Saalfeld[1], Albert Cardona[1,c]

September 10, 2015

*: equal contribution.

c: corresponding author: cardonaa@janelia.hhmi.org.

1: HHMI Janelia Research Campus, Ashburn, VA, 20147. USA.

2: Institute of Neuroinformatics, University of Zurich and ETH Zurich, Switzerland.

Abstract

Large-scale neuronal circuit mapping using electron microscopy demands laborious proofreading by humans who resolve local ambiguities with larger contextual cues or by reconciling multiple independent reconstructions. We developed a new method that empowers expert neuroanatomists to apply quantitative arbor and network context to proofread and reconstruct neurons and circuits. We implemented our method in the web application CATMAID, supporting a group of collaborators to concurrently reconstruct neurons in the same circuit. We measured the neuroanatomical underpinnings of circuit connectivity in *Drosophila* neurons. We found that across life stages and cell types, synaptic inputs were preferentially located on spine-like microtubule-free branches, "twigs", while synaptic outputs were typically on microtubule-containing "backbone". The differential size and tortuosity of small twigs and rigid backbones was reflected in reconstruction errors, with nearly all errors being omission or truncation of twigs. The combination of redundant twig connectivity and low backbone error rates allows robust mapping of *Drosophila* circuits without time-consuming independent reconstructions. As a demonstration, we mapped a large sensorimotor circuit in the larva. We found anatomical pathways for proprioceptive feedback into motor circuits and applied novel methods of representing neuroanatomical compartments to describe their detailed structure. Our work suggests avenues for incorporating neuroanatomy into machine-learning approaches to connectomics and reveals the largely unknown circuitry of larval locomotion.

Introduction

Mapping circuits from electron microscopy (EM) volumes is hard (Helmstaedter, 2013). Manually working through large volumes is slow and prone to attentional errors (Kreshuk et al., 2011; Helmstaedter et al., 2011). Combining multiple independent reconstructions of the same neuron can reduce errors (Helmstaedter et al., 2011; Kim et al., 2014) at the cost of multiplying the required labor. Current computational approaches operate only with "local" information, i.e. the EM micrographs and algorithmically detected fine structures such as cell membranes and mitochondria. They are therefore sensitive to noise (Jain et al., 2010), particularly in anisotropic EM data where the smallest neurites may be thinner than the thickness of individual serial sections (Veeraraghavan et al., 2010; Helmstaedter, 2013). Machine-generated neuron reconstructions are therefore proof-read by humans (Chklovskii et al., 2010; Haehn

et al., 2014).

Experts are able to resolve ambiguities that amateurs and current algorithmic approaches cannot by using large-scale features about neurons to inform decisions made at the level of nanometer-scale image data. In *Drosophila*, where neurons are highly stereotyped, large branches in an EM reconstruction of a given cell can be confirmed by comparing the observed anatomy to that of homologous cells from light microscopy data or other reconstructions (Takemura et al., 2013; Ohyama et al., 2015). This suggests that one way to improve the toolkit for neuron reconstruction and circuit mapping is to facilitate the application of cell- and circuit-level features for disambiguation at noisy locations on EM volumes.

Crucially, different errors do not alter the wiring diagram equally. Missing small dendrites can be acceptable; useful and reproducible wiring diagrams can be created even when omiting 56% of all postsynaptic sites (Takemura et al., 2013) while missing a single large branch hosting all the synapses in one neuropil region could omit connectivity to entire populations of partners. Prioritizing proofreading time towards the most impactful errors improves reconstruction efficiency (Plaza et al., 2012; Kim et al., 2014).

However, to understand the impact of reconstruction errors in *Drosophila*, we need to understand the relationship between circuitry and anatomy. Mesoscale anatomy, in particular the placement of large branches, is well understood as a key component of circuit structure (Zlatić et al., 2003, 2009; Wu et al., 2011; Couton et al., 2015). Similarly, the connectivity graph of a stereotyped circuit can itself relate back to anatomy by consideration of the location of the synaptic sites between pairs of neurons. However, relatively little is known about the smallest scales of synaptic connectivity, the distribution of individual synapses on a neuron. Microtubule-free and actin-rich structures have been identified as sites of excitatory input in the adult *Drosophila* visual system (Scott et al., 2003; Leiss et al., 2009), but questions remain about how ubiquitous these are in the nervous system.

Here, we describe a collection of quantitative anatomical and connectivity features across scales, from fine dendritic branches to multi-neuron graphs, and a method for using them to swiftly and accurately map the *Drosophila* nervous system. We implemented the calculation and visualization of such features on-demand as an extension of the web-based large image data viewer CATMAID (Saalfeld et al., 2009). Because the detection of high-impact errors can occur concurrently with reconstruction via interactive analysis, our tool removes the need for time-consuming repeated reconstructions (Helmstaedter et al., 2013; Kim et al., 2014) and enables synergy among collaborating expert neuroanatomists.

We demonstrate our methods by mapping a sensorimotor circuit in the *Drosophila* larva from proprioceptive sensory neurons to motoneurons.

Results

Collaborative circuit mapping

We extended the web-based image data viewer CATMAID (Saalfeld et al., 2009) to enable a geographically distributed group of researchers to map neuronal circuitry. A neuron is reconstructed with a skeleton, a directed tree graph with one or more nodes in every cross-section of neurite in an EM volume (Helmstaedter et al., 2011; Cardona et al., 2012). Where possible, we root skeletons at the soma to model the anatomical notions of proximal and distal in the data structure. Synapses (see Supplemental Fig. 1) are annotated as a relation from a node on the presynaptic neuron skeleton to an intermediate "connector node" and then to a node of a postsynaptic neuron skeleton. To express the polyadic nature of insect synapses (Meinertzhagen and O'Neil, 1991), connector nodes can have multiple postsynaptic "targets", but only one presynaptic "source". Reconstructions are immediately synchronized across all users to avoid duplicate or conflicting work, and to take advantage of existing reconstructions to aid further reconstruction and circuit discovery. For further details, see Supplemental Text.

As a case study of our method, we focused on sensorimotor circuits in an abdominal segment of the first instar *Drosophila* larval central nervous system (CNS) using an EM volume covering one and a half abdominal segments (Supplemental Fig. 2; previously used in Ohyama et al. (2015)). In total for this work, we reconstructed 425 neuronal arbors spanning 51.8 millimeters of cable, with 24,068 presynaptic and 50,927 postsynaptic relations, (see Supplemental Text for details). Reconstruction time was 469 hours for reconstruction with synapse annotations plus 240 hours for review (see below), for an average rate of $\sim$73 microns of proofread arbor with synapses per hour.

Microtubule-free twigs are the principal site of synaptic input

To be able to use neuronal anatomy to guide circuit reconstruction, it was crucial to better understand the distribution of synaptic input onto *Drosophila* neurons. We started by looking in detail at the relationship between the microtubule cytoskeleton (Supplemental Fig. 3) and synaptic inputs in EM reconstruc-

tions of neurons from different regions of the nervous system and life stages. For a diverse collection of neurons, we marked all locations where the arbor continued distal to a microtubule-containing process (Fig. 1a). We call such a terminal branch a "twig". By definition, all twigs have their base on a microtubule-containing backbone shaft. Following the classification in Leiss *et al.* (Leiss et al., 2009), a spine is a twig with a maximal depth of less than 3 µm and that is not a presynaptic varicosity (Fig. 1b).

We found twigs in all neurons investigated, across multiple CNS regions and life stages of *Drosophila*, and in all cases they were the dominant sites of synaptic input (Fig. 1c–g). We first considered motoneurons aCC and RP2 (Landgraf et al., 1997), which have functional and structural similarities to vertebrate neurons (Sánchez-Soriano et al., 2005; Nicolaï et al., 2010; Günay et al., 2015). In the first instar CNS, we find aCC and RP2 have numerous twigs, which together host more than 80% of their total number of postsynaptic sites (Fig. 1c). We found a similar majority of inputs onto twigs in three hemisegmental pairs of premotor interneurons (Fig. 1d) and brain neurons (Ohyama et al., 2015) in the first instar (Fig. 1e). We tested whether the observed distribution of postsynaptic sites onto twigs generalizes across larval stages by comparing a somatosensory interneuron in the first instar to its homologue in late third instar (Fig. 1f). At both life stages, we find more than 80% of inputs were onto twigs, suggesting that twigs are not a temporary developmental structure. In the adult fly, light microscopy-level analysis of lobula plate tangential cells of the visual system suggests a similar distribution of postsynaptic sites onto twigs (Leiss et al., 2009; Scott et al., 2003). We annotated EM skeletonizations of medullar Tm3 neurons reconstructed by Takemura et al. (2013) in the adult visual system neuropil and found that nearly all their inputs were onto twigs (Fig. 1g). Our findings suggest that microtubule-free twigs are a general feature of *Drosophila* neurons and constitute the primary anatomical location of synaptic input. Spine-like twigs are found in all cell types, but host a variable, typically non-majority, amount of synaptic input (Fig. 1c–g). We consider all twigs for the remainder of our analysis.

Distribution of inputs onto motoneuron dendrites

For a given presynaptic partner, a postsynaptic neuron could concentrate its input synapses onto a single region or distribute them widely. The spatial distribution of synaptic inputs has implications for dendritic processing (Polsky et al., 2004), developmental robustness (Couton et al., 2015), and as we show, reconstruction accuracy.

To study the relationship between presynaptic neurons and the anatomical locations of post-

synaptic sites, we reconstructed all neurons synaptically connected to motoneurons aCC and RP2 in the third abdominal segment of a first instar larva (Fig. 2a-d). A dynamically-generated and interactive table of synaptic connectivity in CATMAID enabled users to systematically trace all connected arbors. We found 198 identifiable neurons (Supplemental Fig. 4) and named them according to a developmental lineage-based nomenclature (Ohyama et al., 2015) and classified 107 other arbors spanning the full segment into eight distinct intersegmental bundles (Supplemental Fig. 5). Motoneurons each received between 1 and 28 synaptic inputs from individual presynaptic neurons, with a maximum of 7.3% of all inputs coming from a single neuron (Fig. 2e). The fraction of synapses accounted for by their presynaptic partners (rank-ordered by number of synapses) is well-fit by an exponential survival function, with a decay indicating that approximately the top 22 presynaptic partners of one motor neuron contribute 63% of all its synaptic inputs (Fig. 2f).

We next asked how the synaptic input onto aCC and RP2 is distributed across independent twigs. Most individual twigs were small, with the median twig measuring 1 µm in cable and hosting 1 postsynaptic site. The largest typical twig had 16 µm of cable and 20 postsynaptic sites (Fig. 2g, but see Supplemental Text for discussion of three larger outliers). We find that presynaptic neurons connect to between 0 (backbone only) and 13 twigs, with nearly all connections with 3 or more synapses per edge being distributed across multiple twigs (Fig. 2h). Similarly, numerically strong edges spanned multiple twigs in the adult visual system Tm3 neurons (Supplemental Fig. 6).

Mitochondria and presynaptic sites

We next looked for ultrastructural features related to presynaptic sites in the six premotor interneurons described in Figure 1d, since the motoneurons had no central synaptic output. We found that, in stark contrast to inputs, presynaptic sites were often associated with microtubules (Supplemental Fig. 7). Approximately 50% of presynaptic sites are located on the backbone and 90% were within 3 µm of it. Since presynaptic sites are metabolically demanding (Laughlin et al., 1998), we also annotated the location of all mitochondria in the neurons (Supplemental Fig. 7). We found that all presynaptic sites were within 3 µm of their closest mitochondrion, though only 47% of neuronal cable was within 3 µm of a mitochondrion. Taken together, this suggests that the internal structure of neurons puts strong limits on the location of presynaptic sites.

Circuitry for proprioceptive feedback into a motor circuit

We next looked at the cell and circuit level for regularities that could inform proofreading. In the *Drosophila* larva, developmentally homologous neurons are strongly stereotyped (Li et al., 2014), making quantitative analysis of their properties useful for identifying irregularities between homologous cells. Most cell types are represented in the fly nervous system by at least one homologous bilateral pair of individual cells. Bilateral homology suggests both arbor morphology and synaptic wiring are mirrored, up to developmental noise (Ohyama et al., 2015). To let morphology guide proofreading, we developed a collection of neuroanatomical measurements that were independent of absolute location. These metrics, combined with 3d visualization, quickly summarize the structure of complex neurons to help identify and localize inconsistencies (Supplemental Fig. 8) which could result from true developmental differences (Supplemental Fig. 9) or reconstruction errors.

As a case study, we applied our tools to describe a complete sensorimotor circuit. During forward crawling, a peristaltic wave of muscle contraction travels from posterior to anterior segments (Hughes and Thomas, 2007; Heckscher et al., 2012). Signals from the segmentally repeated proprioceptive neurons dbd have been suggested to act via a polysynaptic inhibitory pathway to stop motoneuron activity after successful contraction of a given segment (Hughes and Thomas, 2007). To find pathways between proprioceptive and motor neurons, we further reconstructed axons for proprioceptive sensory neurons dbd, vbd, dmd1, ddaD, and ddaD (Hughes and Thomas, 2007; Grueber et al., 2007). Because of its implication in proprioceptive feedback (Hughes and Thomas, 2007), we further reconstructed all partners of the left and right dbd (Fig. 3a).

Using a graph search within CATMAID, we identified all 1–3 hop pathways from dbd to motoneuron RP2. Comparison of the identifiable intermediate neurons revealed five pairs of homologous neurons with consistent shape, connectivity, and basic morphological quantities (Fig. 3b,c). Inconsistencies in any property led to further review to determine if they were due to reconstruction error, true developmental variability, or limitations of the raw data. For example, one strong inconsistency in this network, the connection from A02l to A31k (Fig. 3b), was due to the expected synapse locations being outside the imaged volume on one side but not the other (Supplemental Fig. 2).

The five pairs of identified neurons could also be matched to light-level images of neurons identified through sparse stochastic labeling (Nern et al., 2015) of neurons within a GAL4 expression pattern (Fig. 3c). Of these, two directly premotor interneurons (A27j and A31k) are immunoreactive to

anti-GABA (Supplemental Fig. 10), whereas the others, all from A02 lineage, are members of the glutamatergic neuron class described in Kohsaka et al. (2014). These novel, putatively inhibitory sensorimotor pathways are well-positioned to mediate a hypothesized "mission accomplished" signal (Hughes and Thomas, 2007). This map also could explain why genetic silencing of A02 neurons was shown to slow peristalsis (Kohsaka et al., 2014), as doing so removes a major channel for proprioceptive feedback.

Anatomically enriched wiring diagrams reveal propriomotor circuit motifs

The physiology of synaptic input and output can differ across neuronal compartments. For example, presynaptic inhibition is important for gain control in fly sensory circuits in a fundamentally distinct manner than dendritic inhibition (Clarac and Cattaert, 1996). This suggests that connectivity can be stereotyped at the compartmental level and therefore useful for proofreading. We thus sought a graph representation of a circuit that could faithfully distinguish distinct types of connections (Fig. 3d-g).

In *Drosophila*, many neuronal cell types have distinct input and output compartments, while a few have entirely intermingled inputs and outputs. Our approach assumes that the neuron can be split into distinct compartments, and at the end checks to see if the split was successful. First we calculate all paths along the skeleton from each of the neuron's input synapses to each of its output synapses and for each node of the skeleton compute the number of centripetal (towards soma) and centrifugal (away from soma) paths that pass through it (Fig. 3d). This quantity, which we call "synapse flow centrality" (SFC), is analogous to a synapse-specific version of betweenness centrality (Newman, 2010). For most neuronal arbors, we find that the most proximal skeleton node with the highest centrifugal SFC corresponds to an intuitive generalization of the locations of spike initiation zones in known polarized neurons of *Drosophila* (Gouwens and Wilson, 2009; Günay et al., 2015) and other insects (Gabbiani et al., 2002).

We quantify how completely input and output are separated on a neuron with a "segregation index", an entropic measure of the amount of input/output mixing in each compartment, normalized by that of the whole arbor (see Methods and Supplemental Fig. 11). A very low segregation index means that pre- and post-synaptic sites are intermingled and an axon/dendrite compartmentalization is inappropriate. Using this approach, we classified all identifiable neurons found in both the left and right hemisegments of the proprio-motor circuitry described above. Of the 3834 synapses between these cells, we found 79% were axo-dendritic (3033), 11% axo-axonic (424), 9% dendro-dendritic (334) and 1%

dendro-axonic (43).

We consider two examples of how compartment-enriched graphs add important anatomical detail to small microcircuits. First, we analyzed how different proprioceptive inputs converge onto motoneuron RP2 (Fig. 3e–h). By splitting interneuron A02b into axon and dendrite we observed that its dendrites receive bilateral proprioceptive input, while its axon synapses both onto the ipsilateral RP2 and axo-axonically onto its strong premotor partner, A03a1 in both hemisegments (Fig. 3f). In contrast, while dbd only connects indirectly with A02b (Fig. 3b), it synapses exclusively ipsilaterally and axo-axonicaly onto A03a1 (Fig. 3f). This suggests that the role of dbd in modulating motor patterns could be qualitatively different than the other proprioceptive sensory neurons, since its direct pathways are typically longer or involve connections types other than axo-dendritic.

Second, we analyzed interactions between the premotor neurons of aCC and RP2 (Fig. 3g,h) . We found that a neuron presynaptic to the aCC motoneuron on both sides targets dendro-dendritically a pre-RP2 neuron (A27h), potentially coordinating the joint excitation of their targets (Fig. 3g). We also found a premotor interneuron (A27e) with reciprocal connections with a GABAergic premotor interneuron (A27j; see Supplemental Fig. 10) that receives convergent inputs from dorsal proprioceptive neurons (dmd1, ddaD, ddaE; Fig. 3g,h). This suggests that A27j might not only act as an inhibitory premotor input in response to proprioceptive activity, but also have subtler modulatory effects onto other sources of motor input.

Specific connections can also be allocated to specific regions of a neuronal arbor, which could be used to localize proofreading guided by inconsistencies in connectivity. We thus extended the concept of splitting a neuron into two regions to an arbitrary number, by defining a region as a cluster of synapses near each other along the arbor cable (see Methods). As an example, we consider the axon terminal of dbd, which enters at the interface between two segments and extends symmetric arbors towards the anterior and posterior segments (Fig. 3i). The synapses form multiple well-separated clusters that we can visualize as a group of nodes, revealing that the anterior and posterior branches synapse onto homologous interneurons (A08a) for their respective segments (Fig. 3j). This pattern suggests that each A08a cell gets convergent input from the dbd of two consecutive segments, which could reflect that adjacent pairs of segments move together during locomotion (Heckscher et al., 2012).

Proofreading and error correction

Based on the spectrum of features described above, we developed a two step iterative method of proofreading after an initial reconstruction. An initial systematic review consists in traversing a whole arbor from distal to proximal to follow the expected gradual changes in anatomical properties (e.g. caliber tapering and cytoskeletal changes from microtubule-free to increasing number of microtubules). This review mode helps discover attentional errors or anatomical inconsistencies, such as non-contiguous microtubule cytoskeleton. Next, we reconstruct a second instance of homologous circuitry and use the high-level quantitative anatomical and connectivity measurements to highlight inconsistencies, which are then subjected to focused review (see Supplemental Text for details). This review mode helps ensure that the broad structure of the neuron is consistent and no high-impact large branches are missed.

Importantly, any error correction is performed on the basis of local information, as contained within the EM images (e.g. microtubules, texture, consistency with neighboring neurites, etc. See Supplemental Text). Irregularities in higher level features only serve to guide attention, not determine correctness. The strong stereotypy of *Drosophila* neurons (Cardona et al., 2010; Jenett et al., 2012; Li et al., 2014) enables the use of light microscopy images or contralateral homologs in the EM for overcoming ambiguities in large, microtubule-rich backbone branches, but not for more variable twigs. To do this requires partial annotation of all feasible continuations within the EM volume to ensure that there is only single solution consistent with the stereotypic features (see Supplemental Text). We also annotate ambiguities during reconstruction with a scale from 1 to 5 for further consideration in subsequent quantitative or visual analysis of the arbors and the wiring diagram.

Validation of our iterative circuit mapping method

Our approach to circuit mapping consists of an initial reconstruction, followed by edits by the same or different users during proofreading or incidental discovery of errors during subsequent work. Small arbor pieces, left over from pruning when proofreading other neurons or from explorative tracings, are merged in. We refer to this as "iterative," as compared to consensus methods that combine multiple independent reconstructions (Helmstaedter et al., 2011; Takemura et al., 2013; Kim et al., 2014).

We evaluated the accuracy of our method for *Drosophila* circuits by comparing our results to the those of a consensus approach. We selected six interconnected neurons from the premotor network

for independent reconstruction by four individuals. Each individual skeletonized and reviewed his or her reconstructions and then consensus skeletons were computed for each neuron using RESCOP (Helmstaedter et al., 2011). Both methods resulted in extremely similar arbors, although each method found branches not seen in the other (Fig. 4a, Supplemental Fig. 12). All sites of disagreement between the two methods were validated by an expert to determine a gold-standard morphology. Reconstruction and review of these six neurons in the iterative approach took a total of 26.37 hours, while the redundant method by four people took a total of 107.73 hours, almost exactly four times as long.

Existing consensus approaches only calculate neuronal morphology, not synaptic connectivity. We estimated the connectivity between consensus skeletons by transferring each synapse from each independent skeleton to the consensus, normalized by number of skeletons contributing to the consensus at the location of the synapse. A synapse would have unit weight when annotated at the same location in all independent skeletons.

We found that both methods recover an identical set of edges in the wiring diagram, with similar number of synapses per edge (Fig. 4b,c). We next considered the fine differences between consensus skeletons and skeletons reconstructed with our method. The six gold-standard neurons had a total of 1341 input synapses, with 111 on neurites only present in the consensus skeletons, 229 on neurites only in our method's reconstructions, and 1001 in the arbor found by both. We located 91 missed or incomplete branches (false negatives) in our method, 89 in twigs and 2 in backbones; and 7 incorrect continuations (false positives), 6 in twigs and 1 in backbone. False positives added 30 incorrect postsynaptic inputs in total. Individual missed branches were small in size, complexity, and number of synapses (Fig. 4e–g), with more than 40 missed or truncated twigs having no impact on connectivity (Fig. 4f). The 3 errors in backbones occurred in small distal dendritic shafts containing one single microtubule, resulting in 7 missed and 4 false postsynaptic sites. Error rates for synaptic output were even lower. The gold-standard neurons had a total of 510 presynaptic sites, of which 509 were found by our iterative reconstructions.

Our data suggest that anatomical structure strongly influences the rate of reconstruction errors in our iterative method. Our total error rate is dominated by false negatives and is much higher for twigs (16.2 µm/error) than for backbone (375.8 µm/error). While attentional errors seemed to dominate missed branches, data ambiguities were often associated with backbone errors. One backbone false merge happened across two adjacent sections in poor registration with one another, while an er-

roneous truncation occurred across a section where electron-dense precipitate occluded the neurite and its surrounding area.

Estimating errors in a reconstructed wiring diagram

Neuroanatomy strongly constrains the impact of a typical error on the wiring diagram because, as shown above, the most likely error is to miss a twig and an individual twig hosts few or no synapses.

To estimate the probability of omitting a true edge in the wiring diagram, we analyzed the distribution of synaptic contacts across twigs as a function of the total number of synapses per edge. Edges comprising multiple synaptic contacts were found to be distributed across multiple twigs (Fig. 2h). With the RESCOP-based validation we found that our method identified 88% (672/761) of twigs, containing 91.7% of synapses (1230/1341). From these two observations, we estimated the probability of completely missing a true edge as a function of the number of morphological synapses per edge. We found that our method recovers more than 99% of the wiring diagram edges that have at least 3 synapses (Fig. 5a), assuming twigs are missed uniformly at random (see Supplemental Fig 13).

In the larva, we are primarily interested in the most reliable edges between cell types, as those are most likely to generalize across individual animals. Moreover, we are concerned less about adding extra synapses to true connections and more about adding false edges that would be interpreted as pathways that are not actually present. To estimate the likelihood of introducing a false edge between cell types not just once, but twice (e.g. in a left and right pair of homologs), we simulated adding false twigs to a neuron. The probability of adding a false edge depends both on the probability of adding a false twig (observed false positive error rate: 7 errors in 605 twigs) and the number of nearby but unconnected neurons with presynaptic sites. This will vary depending on the circuit in question. For example, a neuropile with all-to-all connectivity will have no opportunity for false positive edges, while in an array of rigorously separated labeled lines any false positive synapse would be a false positive edge. Further, larger neurons offer more opportunities for false positives than smaller neurons.

For a concrete and realistic example, we consider the motoneuron RP2 (a large neuron). We estimated the number of proximate but unconnected neurons by considering all axons presynaptic to all motoneuron dendritic fields that overlap RP2's dendrites (Fig. 5b). We assume that a false positive reconstruction error distributes m synapses across all available axons at random. Even if we assume that m is always among the largest observed ($m = 20$), our model suggests that for the RP2 wiring

diagram we can trust symmetric connections of at least 2 synapses (Fig. 5c). We further note that the small size of individual twigs and the ability in CATMAID to jump directly to the image data associated with synapses comprising an edge make review of a suspect false positive edge extremely fast, on the order of seconds.

Since most errors were of omission and took the form of cropped twigs, we also measured the impact of omitting the distal ends of twigs. Considering again aCC and RP2, we looked at the connectivity observed by considering only synapses located at a given depth into the twig relative to its base on the backbone (Fig. 6a). With all twigs cropped to zero depth, only inputs onto the backbone remain. More than 90% of postsynaptic sites lay within 5 µm of the backbone (Fig. 6b). We find that the first 2 µm already finds at least two synapses between ∼90% of the most connected partners. The first 4 µm similarly detects ∼90% of all partners with 2 or more synapses and 27/28 pairs of homologous edges (Fig. 6b). These results indicate that, given the observed distribution of synapses over multiple twigs (Fig. 2h), edges with many synapses are robust to errors of omission such as truncated twigs. Considering the marginal time involved in reconstructing the full extent of twigs (Fig. 6c), this robustness could be intentionally exploited towards discovering strong synaptic partners in a time-efficient manner.

Discussion

Neuroanatomy as the foundation for circuit mapping

Neurons are highly structured cells. A human expert's success at circuit mapping from EM volumes stems from the ability to use this structure and apply cell and circuit-level context to interpret nanometer-scale image data. Here, we presented our approach to circuit mapping in EM by building tools in CATMAID that ease and emphasize the use of high level features concurrent with image-level reconstruction. In addition to using existing measures of neuroanatomy and graph connectivity, we devised novel algorithms to quantify the distribution of synapses across neurons, which we applied to both quantitative neuroanatomy and to enrich wiring diagrams.

Central to our approach is the observation that *Drosophila* neurons contain a contiguous microtubule-rich backbone and numerous small microtubule-free distal twigs. We found that small twigs are the primary site of synaptic input for *Drosophila* neurons and that numerically strong connections between neurons are typically spread across many distinct twigs. If, contrary to observations, neurons were to

only contact each other via a single twig that hosts many postsynaptic sites, then this connection would be fragile with respect to developmental noise (Couton et al., 2015). In contrast, backbones define the spatial extent and stereotyped shape of a neuron, and we found that most presynaptic sites are located on or very near the backbone's microtubules and mitochondria. Our findings are consistent with the notion that metabolic needs and microtubule-based trafficking are limiting factors for the structure of synaptic output.

These different biological requirements for different neuronal compartments are reflected in the rate of reconstruction errors. The large calibers and relatively gradual turns associated with microtubules made errors on backbone less frequent by a factor of nearly 20 relative to on smaller and tortuous twigs. However, we propose that the circuit's resilience to developmental noise, achieved in part by connecting via multiple twigs, underlies the resilience of wiring diagrams to the omission of small dendritic branches, the most typical error observed both here and in reconstructions in the fly visual system (Takemura et al., 2013).

Irregularities within a cell type guide review toward small fractions of specific neuronal arbors that could be responsible for a potential error. While reconstructing a neuron, a user can quickly pull up its complete anatomy and connectivity to compare to homologous cells or inspect for irregularities and, crucially, return quickly to the locations in the image data necessary to make the appropriate decisions. We find that this smooth flow from image data to high level features and back to image data — without *post hoc* or offline analysis — is possibly the most important feature in our EM reconstruction technique.

Dispensing with repeated reconstruction without reducing accuracy enables our method to support concurrent neuron reconstruction by many collaborators. This setup prevents duplicated work while ensuring that important locations are visited multiple times. For example, synaptic relations are inspected at least twice in different ways, once each from the pre- and postsynaptic side. The presence of existing and correct skeletons in complicated areas, such as registration errors between consecutive sections or gaps, reduces the time necessary for resolving possible ambiguities and effectively provides an extra step of proof-reading by not allowing contradictory reconstructions. Further savings originate in the reuse of data, for example exploratory reconstruction of backbones in search of specific neurons or branches pruned during proofreading are merged into the arbor currently being reconstructed. In summary, in a collaborative environment, the more neurons that are reconstructed, the faster new ones can be added, and the fewer errors existing reconstructions will contain.

Automated methods will be necessary to map circuits with more than a few thousand neurons (Helmstaedter, 2013), but they require extensive proof-reading (Chklovskii et al., 2010; Plaza et al., 2012; Haehn et al., 2014). Our methods for analysis of arbor morphology, synaptic distribution and circuit structure and reliability, and their application in proof-reading, apply equally to manually and automatically reconstructed neurons. Neuroanatomical measurements suggest mixed strategies for leveraging both automated algorithms and human effort. For example, mitochondria can be reliably located automatically (Lucchi et al., 2011; Funke et al., 2014) which, together with our finding of a distance constraint between mitochondria and presynaptic sites, could assist in automated synapse detection (Kreshuk et al., 2011; Becker et al., 2012; Kreshuk et al., 2014). Similarly, the properties of neuronal backbone and twigs suggest that algorithms for the automatic detection of microtubules in serial section EM would be a profitable source of constraints for automated reconstruction of neurites across consecutive sections (Vazquez-Reina et al., 2011; Funke et al., 2012). While we only considered the relationship between error rate and the presence or absence of microtubules, with the use of automated detection methods it will be important to look at more detailed measures of arbors such as the number of microtubules, curvature, or caliber.

We expect that the fundamental principles described here will hold for regions of the nervous system of the adult fly other than the optic neuropils, although the exact properties, e.g. mean twig length, will likely differ across time and cell type. In contrast, applying our method to vertebrate neurons would require extensive adaptation. Connectivity and morphology are less stereotyped, making it impossible to build strong expectations from a single example of a cell type. However, given careful measurement of cell types, such as the distribution of branches or synapses in a given retinal layer already used in retinal reconstructions (Helmstaedter et al., 2013; Kim et al., 2014; Seung and Sümbül, 2015) or automated cell type discovery using anatomy and connectivity (Jonas and Kording, 2015), similar neuroanatomically-guided reconstruction and proofreading could inform methods for vertebrate connectomics. The key challenge will be to find, as we did here for *Drosophila* neurons, measurable quantities that are most informative for discovering those errors that have the greatest impact on the interpretation of wiring diagrams.

Sensorimotor circuitry

Larval locomotion, like many motor patterns, results from rhythmic activation of motoneurons (Heckscher et al., 2012), but few central components of the underlying premotor circuitry had been identified (Kohsaka et al., 2014; Couton et al., 2015). Our reconstruction of propriomotor circuitry revealed a complex network comprised of numerous cell types, including a subset of those previously described (Kohsaka et al., 2014). We identified a rich collection of local neurons, including neurons anatomically well-suited to provide common drive to synergistic muscles (Schaefer et al., 2010) and thus likely a key motor network components. Using anatomically faithful simplifications of neuronal structure, we found several premotor microcircuits employing dendro-dendritic and axo-axonic synapses in parallel with conventional axo-dendritic synaptic connections. For example we found a GABAergic input pre- and post-synaptic to motoneuron input, a motif also observed in mammalian motor circuits (Fyffe and Light, 1984).

Although the motor system is rhythmically active in the absence of sensory input (Suster and Bate, 2002), proprioceptive sensory feedback is required for natural coordination and timing (Hughes and Thomas, 2007; Song et al., 2007). We found diverse and complex circuitry for relaying proprioceptive information, including GABAergic and glutamatergic neurons directly relaying proprioceptive input to motoneurons. This motif is well-posed to provide an inhibitory "mission accomplished" signal to surpress motoneuron activity after a successful contraction during forward locomotion (Hughes and Thomas, 2007). However, we also observed that many synaptic partners of dbd were themselves presynaptic to neurons downstream of the other proprioceptive axons, suggesting other, more complex roles for proprioceptive feedback in modulating motor activity. Surprisingly, the axon terminals of proprioceptive neurons themselves almost entirely lacked presynaptic input. This suggests that proprioceptive input is privileged by the larval nervous system and not under fast, dynamic modulation by central circuitry (Clarac and Cattaert, 1996), unlike proprioceptive afferents in the locust leg (Burrows and Matheson, 1994) and other somatosensory modalities in the larva (Ohyama et al., 2015).

Wiring diagrams have been deemed necessary, yet not sufficient, for understanding neural circuits (Bargmann, 2012) and a fast approach for discarding hypotheses of circuit function (Denk et al., 2012; Takemura et al., 2013). The neuronal wiring reconstructed here offers insights into the structure of an insect motor circuit and its control by sensory feedback, and serves as a complementary resource for detailed functional studies. With the circuit mapping tools and methods demonstrated here, fast,

accurate, and targeted reconstruction of circuits in *Drosophila* larva (Ohyama et al., 2015) and adult, and other species (e.g. *Platynereis* (Randel et al., 2015)) is possible.

Methods

Serial-section Transmission Electron Microscopy

Wild-type *Drosophila* first instar larval central nervous systems were manually dissected by mechanical separation of the anterior tip of the larva from the posterior portion in PBS, and immediately transferred to 2% glutaraldehyde in 0.1 M Na-cacodylate, pH 7.4 buffer. Samples were post-fixed in 1% OsO_4 in the same buffer and stained *en bloc* with 1% aqueous uranyl acetate before subsequent dehydration in ethanol and propylene oxide, and embedding in Epon. Serial 45 nm sections were cut with a Leica UC6 ultramicrotome using a Diatome diamond knife, and picked up on Synaptek slot grids with Pioloform support films. Sections were stained with uranyl acetate followed by Sato's lead (Sato, 1968). Sections were then imaged at 4.4 nm x 4.4 nm resolution using Leginon (Suloway et al., 2005) to drive an FEI Tecnai 20 transmission electron microscope. The resulting 77,000 image tiles were contrast-corrected, montaged and registered with TrakEM2 (Cardona et al., 2012) using the nonlinear elastic method (Saalfeld et al., 2012). The generated data volume of $22775 \times 18326 \times 462$ voxels corresponds to a volume of $91 \times 73 \times 21 \mu m^3$. The data set covers approximately the posterior half of abdominal segment A2, and a nearly complete abdominal segment A3.

Immunolabelings and light microscopy

Central nervous systems of *Drosophila* larva were dissected in PBG (10% NGS [Normal Goat Serum] in 1% PBS) with tweezers under a scope and fixed with 4% paraformaldehyde in 1% PBS for 30 min, washed 3×10 minutes in PBT (1% Triton-X100 in 1% PBS), blocked for 1 h in PBG, then washed 3×10 minutes in PBGT (1% Triton-X100 in PBG), and incubated with primary antibodies (rabbit anti-GABA: Sigma A2053 at 1/1000; chick anti-GFP: Abcam ab13970 at 1/2000) in PBGT for 24 h at 4^oC on small Eppendorf tubes laid on a gentle horizontal shaker. They were then washed 4×15 min in PBT, and incubated with secondary antibodies (goat anti-chick 488: Invitrogen, at 1/500; goat anti-rabbit 568: Invitrogen, at 1/500) in PBGT at 4^oC in Eppendorf tubes wrapped in aluminum foil on a horizontal

shaker for 24 h, subsequently washed in PBT 4 × 15 min (wrapped in foil), and mounted on poly-lysine coated glass slides. Then samples were dehydrated by dipping the slides in an alcohol series (30%, 50%, 70%, 90% in distilled water, then twice 100%) and then in 100% xylene (3 times) using Columbia glass jars with slits for slides; then mounted on glass slides in DePeX (Li et al., 2014) using spacer coverslips on the sides. Glass slides were left to dry in a large Petri dish with a lid, wrapped in foil and at 4^{o}C for 3 days. Imaging was done with a Zeiss 710 confocal laser-scanning microscope. Positive immunoreactivity was confirmed by consistent labeling across multiple GFP-labeled cells per imaged nervous system in two or more nervous systems.

Synapse clustering algorithm

In order to associate synaptic connectivity not to whole neurons, but to regions of neurons, we adopt an approach where we cluster nearby synapses. Mean shift clustering has been shown to be an effective approach to finding synapse clusters in 3d without assuming a particular number of groups a priori (Binzegger et al., 2007). This approach involves convolving synapse locations with a Gaussian kernel to estimate the density of synapses in space. A cluster is then the set of synapses for which, starting at their location, gradient ascent reaches the same density peak. However, locations on one neuron that are close in space can be very far apart along the neuron. Here, instead of considering the density of a neuron's synapses in 3d space, we use a similar procedure to estimate the density of synapses at every point on the arbor (following the cable) and define synapse clusters in the same manner. The only parameter in both approaches is the width of the Gaussian kernel, a physically meaningful parameter.

For these purposes, we define the skeletonization of a neuron to be a graph with a set of nodes $\mathcal{N}$ with locations $\mathbf{x_i}$ for $i \in \mathcal{N}$ and edges $\mathcal{E}$. Because the neuron's graph is tree-like, there is a unique non-overlapping path on the graph between any two points $i, j \in \mathcal{N}$ with distance δ_{ij}. All synapses (both inputs and outputs) associated with the neuron are represented by the set of their associated nodes, $\mathcal{S} \subset \mathcal{N}$, noting that the same node can be associated with multiple synapses and thus appear multiple times in $\mathcal{S}$. For every node in the neuron graph, we compute the synapse density

$$d(i) = \sum_{j \in \mathcal{S}} \exp\left(-\frac{\delta_{ij}^2}{2\lambda^2}\right)$$

where λ is a bandwidth parameter that effectively determines the size of clusters, and presynaptic sites

of polyadic synapses are counted as many times as they have postsynaptic partners. Note that due to branches, a single synapse close to a branch point may contribute more total density than one that is very distant, reflecting its greater within-graph proximity to more of the arbor. We then look for all maxima in the synapse density and the basins of attraction that flow to them via gradient ascent (i.e. starting at a given node, moving along the maximally positive difference in density between adjacent nodes). A cluster of synapses is then all synapses associated with nodes found within a single basin of attraction of the density function. For neurons found in the 1st instar larva, with ≈500–2000 µm of cable, bandwidths around 8–30 µm provide clusterings that match the subjective description of either "dendritic arbor" or "axon". Smaller bandwidth values result in more granular breakdowns of dendritic and axonal trees (e.g. dbd axons in Figure 5g).

Synapse flow centrality of segments of a neuronal arbor

We define synapse flow centrality (SFC) as the number of possible paths between input synapses and output synapses at each point in the arbor. We compute the flow centrality in three flavors: (1) centrifugal, which counts paths from proximal inputs to distal outputs; (2) centripetal, which counts paths from distal inputs to proximal outputs; and (3) the sum of both.

We use flow centrality for four purposes. First, to split an arbor into axon and dendrite at the maximum centrifugal SFC, which is a preliminary step for computing the segregation index, for expressing all kinds of edges (e.g. axo-axonic, dendro-dendritic) in the wiring diagram, or for rendering the arbor in 3d with differently colored regions. Second, to quantitatively estimate the cable distance between the axon terminals and dendritic arbor by measuring the amount of cable with the maximum centrifugal SFC value. Third, to measure the cable length of the main dendritic shafts using centripetal SFC, which applies only to insect neurons with at least one output synapse in their dendritic arbor. And fourth, to weigh the color of each skeleton node in a 3d view, providing a characteristic signature of the arbor that enables subjective evaluation of its identity.

Segregation index: a measure of synaptic sign distribution in a neuronal arbor

An ideal, textbook neuron has a purely dendritic tree and a purely axonal tree, that is, one tree that only receives inputs and another that only delivers outputs onto other neurons. In reality, dendro-dendritic

and axo-axonic synapses are common. We have observed that homologous neurons (e.g. identifiable neurons in the left and right hemisegments) have a similar synaptic distribution, which differs from that of other neurons. In insects, we find (1) neurons that abide to the canonical model as outlined above—separated input and output trees—such as motor neurons and many types of projection neurons; (2) neurons that consist of a unique tree with mixed inputs and outputs (some of these neurons have been reported as non-spiking (Burrows, 1992)); and (3) everything in between.

Having clustered synapses into groups (either by synapse clustering or by splitting the neuron by centrifugal synapse flow centrality), we ask how neuronal inputs and outputs are distributed among the clusters. If the clustering can adequately separate axon from dendrite, then a highly polar neuron will have most of its outputs on the "axon" cluster and most of its inputs on the "dendrite" cluster. Motor neurons in the abdominal segments of the *Drosophila* larva are examples of completely polarized neurons. Conversely, highly non-polar neurons will have inputs and outputs distributed homogeneously throughout their arbor. An example of the latter are non-spiking neurons that perform extremely local computations, such as GABAergic antennal lobe interneurons (Wilson and Laurent, 2005).

A measure of synaptic sign distribution in a neuronal arbor has the potential to distinguish similar yet uniquely different neurons, as well as to suggest broad functional roles of the neuron. Here, we describe an algorithm to quantify the degree of segregation between input and outputs in a neuronal arbor.

For each synapse cluster i on a neuron with N_i synapses, compute the fraction p_i that are postsynaptic. We measure the uniformity of the distribution of inputs and outputs within cluster i by computing its entropy, for which we consider synapses as entities with two possible states: input or output. At the limits, when all synapses of the cluster are either inputs or outputs, its entropy is zero. When half of the synapses are inputs and the other half are outputs, the entropy is maximal. The contribution of each cluster i to the total entropy is weighted by its fraction of the total synapses (either inputs or outputs).

The entropy of the input/output distribution for each cluster is then

$$S_i = -\left(p_i \log p_i + (1 - p_i) \log\left(1 - p_i\right)\right).$$

The total entropy for the arbor is then just

$$S = \frac{1}{\sum_i N_i} \sum_i N_i S_i.$$

However, for reference we need to compare this to an unstructured arbor (i.e. non-clustered) with the same total number of inputs and outputs; for this we consider the whole arbor as one cluster, and we compute

$$S_{norm} = \frac{\sum_i p_i N_i}{\sum N_i} \log\left(\frac{\sum_i p_i N_i}{\sum N_i}\right) + (1 - \frac{\sum_i p_i N_i}{\sum N_i}) \log\left(1 - \frac{\sum_i p_i N_i}{\sum N_i}\right)$$

(where $\frac{\sum_i p_i N_i}{\sum N_i}$ is just the total fraction of synapses that are inputs).

We define the segregation index as

$$H = 1 - \frac{S}{S_{norm}}$$

so that $H = 0$ corresponds to a totally unsegregated neuron, while $H = 1$ corresponds to a totally segregated neuron. Note that even a modest amount of mixture (e.g. axo-axonic inputs) corresponds to values near $H = 0.5$–0.6 (Supplemental Fig. 11).

Model for estimating false negatives and false positives in the wiring diagram

To estimate the probability of completely missing an edge as a function of the number of synapses in the edge, we combine the twig distribution with the error rates obtained from multi-user reconstruction. We found that our reconstruction identified 672 out of 761 twigs, giving our method a recall rate for complete twigs of $q = 0.88$. Let the distribution of n_b twigs across edges with m synapses be $p(n_b; m)$. Assuming each branch has a probability q of being correctly observed, the probability of not observing a specific connection across all n_b twigs is $(1 - q)^{n_b}$. The probability of omitting an edge with m synapses is thus given by

$$P_{loss}(m) = \sum_{n_b=1}^{m} p(n_b; m)(1 - q)^{n_b}.$$

In our reconstruction method, we emphasize connections that are found consistently between cells of the same type, typically hemisegmental homologs of a presynaptic and postsynaptic neuron. Using a simple model, we approximate the likelihood of introducing a symmetric, but false, edge between

cell types in our wiring diagram due to reconstruction mistakes. Consider two neurons, $j = 1, 2$, of the same cell type, with the dendrites of each sufficiently close to N axons to physically permit connections. To add an incorrect edge to the connectivity graph and not just reweight an existing one, any added branches must have synapses from otherwise unconnected neurons. Let the number of axons with zero true connectivity be N_0. Assuming symmetry, the number of axons for both neurons should be similar. We then suppose that errors add m synapses to each neuron, with each synapse assigned uniformly at random to axon $i \in (1, 2, ..., N)$, with the final added edge count onto neuron j from axon i given by $k_{i,j}$. For clarity, we order the axons such that $i \leq N_0$ designates an axon with no true connectivity. We then ask what is the probability that both $k_{i,1}, k_{i,2} > k_\theta$ for any $i \leq N_0$.

The parameters of this model will vary depending on the properties of the neuron and neuropil in question. For example, larger neurons will have more opportunities for error than smaller ones, while neurons with more stringent synaptic specificity have more true zero edges than broadly synapsing neurons. To estimate a realistic set of values for the neurons here, we consider the validation data. Because nearly all false positives occur on the terminal arbors, the number of synapses added by error m can be expressed as $m = rL_t\bar{k}$, the product of the rate of incorrect branches per length of twig r, the total length of twigs L_t, and the expected synapses per added twig $\bar{k}$. Based on the independent reconstructions, we estimate r as 6 false positive errors per 1.63×10^3 µm, $\bar{k} = 5$ synapses, and a typical $L_t = 257$ µm, making $m = 5$. Determining N and N_0 is difficult, as it requires knowledge of axons that would not be in the connectivity-driven reconstruction. We estimate reasonable values using the aCC and RP2 network, since aCC dendrites strongly overlap RP2 dendrites, but have several presynaptic neurons not shared with RP2 (Figure 4b). In addition to the axons presynaptic to RP2, we find a mean of $N_0 = 36$ inputs exclusive to aCC, so we estimate $N = 87$. We simulated the 10^6 iterations of the model for k_θ = 1–4. To investigate more extreme errors than the ones measured, we also considered $m = 37$ synapses, the largest twin twig observed across all neurons looked at, and $m = 20$ synapses, a more typical value for the largest twig of a single neuron.

Estimating skeleton reconstruction and review time

Skeletons are chimeras, with multiple contributors creating various parts at different points in time. We estimate the total amount of time spent skeletonizing an arbor—including its synapses—by counting the number of 20-second intervals that contain at least one skeleton node or connector node related to the

skeleton. This approach is robust to the discontinuity in time and authorship of adjacent skeleton nodes, but tends to overestimate slightly reconstruction time, given the contribution of 20-second intervals for single nodes that were created earlier in time as pre- or postsynaptic placeholder skeletons with a single node, and which were subsequently merged into the growing skeleton. If the latter were each counted as contributing 6 seconds only, reconstruction times per skeleton typically shrink between 15 and 25%.

We estimate the time for the systematic review of a neuron similarly, with the added caveat that parts of the same arbor may have been reviewed more than once. We count the number of minutes for which at least one skeleton node was reviewed, for every contributor that reviewed a fraction of the arbor, and then add up all the minutes of each contributor.

CATMAID software

We rewrote and greatly developed the Collaborative Annotation Toolkit for Massive Amounts of Image Data, CATMAID (Saalfeld et al., 2009) (GPL), to implement our methods for neural circuit reconstruction, visualization and analysis, and with a user and group management system with sophisticated permissions for graded access. The toolkit consists of four parts: (1) the client (a web page), and three types of servers, namely (2) an application server based on the Django web framework (`https://www.djangoproject.com`), (3) one or more image volume servers, and (4) an instance of the relational database PostgreSQL (`http://www.postgresql.org`) with all non-image data, which includes metadata such as the spatial information of skeletons, the location of which types of synaptic relations, the text annotations, timestamps and provenance of every action. The original web client accesses, in constant time, arbitrary fields of view of remote stored image volumes. We have greatly extended this capability to include 3-way views (XY, XZ and ZY) and a number of color overlays for multi-channel data such as light-microscopy images or computed derivative data such as membrane probability maps. Analysis of neurons and circuits is performed primarily in the client using the programming language JavaScript, relying on a large number of open source libraries for numerical processing, data management and visualization (D3.js, Numeric Javascript, Cytoscape.js, three.js, jsNetworkX, Raphaël, jQuery, SVGKit). Offline analysis for validation and probability calculations was performed by custom scripts in MATLAB (Mathworks). Documentation and installation instructions are available at `http://catmaid.org`.

Preparation of EM images for CATMAID

For display in CATMAID, we Gaussian-smoothed montages of registered EM images (sigma=0.7 pixels, sufficient to remove high-frequency noise to increase the effectiveness of JPEG compression without sacrificing signal) and then generated an image pyramid with five zoom levels and diced it to collections of 256×256 pixel tiles (512×512 and larger can work better for fast Internet connections), stored in JPEG format (75% compression and stripped of headers with jpeg-optim). This approach reduced data storage from over 700 to 90 gigabytes, which were served from a fast seek time solid-state hard drive.

Server and database configuration

We setup a single server machine (Intel Xeon X5660 with 12 cores, 48 GB of RAM, 10 Gb network card) running Ubuntu 12.04 to host the PostgreSQL database, the image server and the Django server. LDAP id caching was enabled for best performance. Images were stored on high-performance solid-state drives mounted with noatime flag or as read-only, and served via proxy with in-RAM varnishd for caching. The database was configured with large shared buffers (4 GB) and autovacuum on (naptime: 8642 min; scale factor 0.4; analyze scale factor 0.2; cost delay -1; cost limit -1) for optimal performance. We chose to serve pages with Nginx, running 8 processes, with epoll on, 768 worker connections, disabled logs and gzip on (except for JPEG image tiles) for best performance, and with public caching and no-expire settings for images. Django was run via Gunicorn with python 2.7 using 8 processes.

Acknowledgements

We thank James Truman for flp-outs of GMR GAL4 lines; Ingrid Andrade and Javier Valdés-Alemán for reconstructing redundant skeletons for RESCOP; John Patton, Ingrid Andrade, Kenny Floria, Alex Berthold van der Bourg, Lukas von Ziegler and Julie Tran for reconstructing about 20% of all arbor cable; Daniel Bonnéry for discussions about statistics; Davi Bock for conceiving the notion of low-confidence edge in an arbor; Goran Cerić, Tom Dolafi and Ken Carlile for IT support; Nic Strausfeld for the term 'twig' and fruitful discussions; Eric Perlman for code and IT tips; Chris Q. Doe, Matthias Landgraf, Akinao Nose, Anna Kreshuck, Fred Hamprecht, Akira Fushiki, Pau Rué, David Wood, Jan Funke, Steve Plaza, Davi Bock and Greg Jefferis for comments; Brett Mensch for discussions; and the HHMI Janelia

Visiting Scientist program (A.C.), the Swiss National Science Foundation grant 31003A_132969 to A.C., HHMI Janelia and the Institute of Neuroinformatics of the University of Zurich and ETH Zurich for funding.

Contributions

AC, CSM designed the study;

AC, RF performed ssTEM;

CSM, AC devised algorithms;

ML, SG, AC, TK, ACh, SS, FM wrote the software;

AC, CSM, SG, ML mapped the circuits;

MZ, FL performed light-microscopy labelings and annotated mitochondria;

CSM, AC, SG analyzed the data;

AC, CSM, SG wrote the paper with input from co-authors.

References

Bargmann, C. (2012). Beyond the connectome: how neuromodulators shape neural circuits. *Bioessays*, 34(6):458–65.

Becker, C., Ali, K., Knott, G., and Fua, P. (2012). Learning context cues for synapse segmentation in em volumes. In Ayache, N., Delingette, H., Golland, P., and Mori, K., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012*, volume 7510 of *Lecture Notes in Computer Science*, pages 585–592. Springer Berlin Heidelberg.

Binzegger, T., Douglas, R. J., and Martin, K. A. (2007). Stereotypical bouton clustering of individual neurons in cat primary visual cortex. *The Journal of Neuroscience*, 27(45):12242–12254.

Burrows, M. (1992). Local circuits for the control of leg movements in an insect. *Trends in neurosciences*, 15(6):226–232.

Burrows, M. and Matheson, T. (1994). A presynaptic gain control mechanism among sensory neurons of a locust leg proprioceptor. *The Journal of neuroscience*, 14(1):272–282.

Cardona, A., Saalfeld, S., Arganda Carreras, I., Pereanu, W., Schindelin, J., and Hartenstein, V. (2010). Identifying neuronal lineages of *Drosophila* by sequence analysis of axon tracts. *J Neurosci*, 30(22):7538–53.

Cardona, A., Schindelin, S. S. J., Arganda Carreras, I., Preibisch, S., Longair, M., Tomancak, P., Hartenstein, V., and Douglas, R. (2012). TrakEM2 Software for Neural Circuit Reconstruction. *PLoS ONE*, 7(6):e38011.

Chklovskii, D., Vitaladevuni, S., and Scheffer, L. (2010). Semi-automated reconstruction of neural circuits using electron microscopy. *Current Opinion in Neurobiology*, 20(5):667–75.

Clarac, F. and Cattaert, D. (1996). Invertebrate presynaptic inhibition and motor control. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, 112(2).

Couton, L., Mauss, A. S., Yunusov, T., Diegelmann, S., Evers, J. F., and Landgraf, M. (2015). Development of connectivity in a motoneuronal network in Drosophila larvae. *Current biology : CB*, 25(5):568–576.

Denk, W., Briggman, K. L., and Helmstaedter, M. (2012). Structural neurobiology: missing link to a mechanistic understanding of neural computation. *Nature Reviews Neuroscience*, 13(5):351–358.

Funke, J., Andres, B., Hamprecht, F., Cardona, A., and Cook, M. (2012). Efficient Automatic 3D-Reconstruction of Branching Neurons from EM Data. *Computer Vision and Pattern Recognition*.

Funke, J., Martel, J., Gerhard, S., Andres, B., Ciresan, D. C., Giusti, A., Gambardella, L. M., Schmidhuber, J., Pfister, H., Cardona, A., and Cook, M. (2014). Candidate Sampling for Neuron Reconstruction from Anisotropic Electron Microscopy Volumes. In *MICCAI Proceedings*.

Fyffe, R. E. W. and Light, A. R. (1984). The ultrastructure of group Ia afferent fiber synapses in the lumbosacral spinal cord of the cat. *Brain research*, 300(2):201–209.

Gabbiani, F., Krapp, H. G., Koch, C., and Laurent, G. (2002). Multiplicative computation in a visual neuron sensitive to looming. *Nature*, 420(6913):320–324.

Gouwens, N. W. and Wilson, R. I. (2009). Signal propagation in *Drosophila* central neurons. *The Journal of Neuroscience*, 29(19):6239–6249.

Grueber, W., Ye, B., Yang, C., Younger, S., Border, K., Jan, L., and Jan, Y. (2007). Projections of *Drosophila* multidendritic neurons in the central nervous system: links with peripheral dendrite morphology. *Development*, 134(1):55–64.

Günay, C., Sieling, F. H., Dharmar, L., Lin, W.-H., Wolfram, V., Marley, R., Baines, R. A., and Prinz, A. A. (2015). Distal spike initiation zone location estimation by morphological simulation of ionic current filtering demonstrated in a novel model of an identified drosophila motoneuron. *PLoS Comput Biol*, 11(5):e1004189.

Haehn, D., Beyer, J., Pfister, H., Knowles-Barley, S., Kasthuri, N., Lichtman, J., and Roberts, M. (2014). Design and evaluation of interactive proofreading tools for connectomics. *IEEE Transactions on Visualization and Computer Graphics*, PP.

Heckscher, E. S., Lockery, S. R., and Doe, C. Q. (2012). Characterization of *Drosophila* larval crawling at the level of organism, segment, and somatic body wall musculature. *The Journal of Neuroscience*, 32(36):12460–12471.

Helmstaedter, M. (2013). Cellular-resolution connectomics: challenges of dense neural circuit reconstruction. *Nature methods*, 10(6):501–507.

Helmstaedter, M., Briggman, K., and Denk, W. (2011). High-accuracy neurite reconstruction for high-throughput neuroanatomy. *Nature Neuroscience*, 14(8):1081–8.

Helmstaedter, M., Briggman, K. L., Turaga, S. C., Jain, V., Seung, H. S., and Denk, W. (2013). Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature*, 500(7461):168–174.

Hughes, C. L. and Thomas, J. B. (2007). A sensory feedback circuit coordinates muscle activity in *Drosophila*. *Mol Cell Neurosci*, 35(2):383–96.

Jain, V., Seung, H., and Turaga, S. (2010). Machines that learn to segment images: a crucial technology for connectomics. *Current Opinion in Neurobiology*, 20(5):653–66.

Jenett, A., Rubin, G. M., Ngo, T.-T., Shepherd, D., Murphy, C., Dionne, H., Pfeiffer, B. D., Cavallaro, A., Hall, D., Jeter, J., et al. (2012). A GAL4-driver line resource for *Drosophila* neurobiology. *Cell reports*, 2(4):991–1001.

Jonas, E. and Kording, K. (2015). Automatic discovery of cell types and microcircuitry from neural connectomics. *eLife*, 4.

Kim, J. S., Greene, M. J., Zlateski, A., Lee, K., Richardson, M., Turaga, S. C., Purcaro, M., Balkam, M., Robinson, A., Behabadi, B. F., Campos, M., Denk, W., Seung, H. S., and The EyeWirers (2014). Space-time wiring specificity supports direction selectivity in the retina. *Nature*, 509(7500):331–336.

Kohsaka, H., Takasu, E., Morimoto, T., and Nose, A. (2014). A group of segmental premotor interneurons regulates the speed of axial locomotion in *Drosophila* larvae. *Current Biology*.

Kreshuk, A., Koethe, U., Pax, E., Bock, D. D., and Hamprecht, F. A. (2014). Automated detection of synapses in serial section transmission electron microscopy image stacks. *PloS one*, 9(2):e87351.

Kreshuk, A., Straehle, C. N., Sommer, C., Koethe, U., Cantoni, M., Knott, G., and Hamprecht, F. A. (2011). Automated detection and segmentation of synaptic contacts in nearly isotropic serial electron microscopy images. *PloS ONE*, 6(10):e24899.

Landgraf, M., Bossing, T., Technau, G., and Bate, M. (1997). The origin, location, and projections of the embryonic abdominal motorneurons of *Drosophila*. *J Neurosci*, 17(24):9642–55.

Laughlin, S. B., de Ruyter van Steveninck, R. R., and Anderson, J. C. (1998). The metabolic cost of neural information. *Nature neuroscience*, 1(1):36–41.

Leiss, F., Grohand, C., Butcherand, N. J., Meinertzhagen, I. A., and Tavosanis, G. (2009). Synaptic organization in the adult *Drosophila* mushroom body calyx. *J Comp Neurol*, 517(6):808–24.

Li, H.-H., Kroll, J. R., Lennox, S. M., Ogundeyi, O., Jeter, J., Depasquale, G., and Truman, J. W. (2014). A GAL4 Driver Resource for Developmental and Behavioral Studies on the Larval CNS of *Drosophila*. *Cell Reports*, 8(3):897–908.

Lucchi, A., Smith, K., Achanta, R., Knott, G., and Fua, P. (2011). Supervoxel-based segmentation of mitochondria in em image stacks with learned shape features. *IEEE transactions on medical imaging*.

Meinertzhagen, I. and O'Neil, S. (1991). Synaptic organization of columnar elements in the lamina of the wild type in *Drosophila melanogaster*. *J Comp Neurol*, 305:232–63.

Nern, A., Pfeiffer, B. D., and Rubin, G. M. (2015). Optimized tools for multicolor stochastic labeling reveal diverse stereotyped cell arrangements in the fly visual system. *Proceedings of the National Academy of Sciences of the United States of America*, 112(22):E2967–76.

Newman, M. (2010). *Networks: An Introduction*. OUP Oxford.

Nicolaï, L. J., Ramaekers, A., Raemaekers, T., Drozdzecki, A., Mauss, A. S., Yan, J., Landgraf, M., Annaert, W., and Hassan, B. A. (2010). Genetically encoded dendritic marker sheds light on neuronal connectivity in *Drosophila*. *Proceedings of the National Academy of Sciences*, 107(47):20553–20558.

Ohyama, T., Schneider-Mizell, C. M., Fetter, R. D., Aleman, J. V., Franconville, R., Rivera-Alba, M., Mensh, B. D., Branson, K. M., Simpson, J. H., Truman, J. W., Cardona, A., and Zlatic, M. (2015). A multilevel multimodal circuit enhances action selection in drosophila. *Nature*.

Plaza, S. M., Scheffer, L. K., and Saunders, M. (2012). Minimizing manual image segmentation turn-around time for neuronal reconstruction by embracing uncertainty. *PLoS ONE*, 7(9):e44448.

Polsky, A., Mel, B. W., and Schiller, J. (2004). Computational subunits in thin dendrites of pyramidal cells. *Nature neuroscience*, 7(6):621–627.

Randel, N., Shahidi, R., Verasztó, C., Bezares-Calderón, L. A., Schmidt, S., and Jékely, G. (2015). Inter-individual stereotypy of the platynereis larval visual connectome. *eLife*, 4:e08069.

Saalfeld, S., Cardona, A., Hartenstein, V., and Tomancak, P. (2009). CATMAID: Collaborative Annotation Toolkit for Massive Amounts of Image Data. *Bioinformatics*, 25(19):1984–1986.

Saalfeld, S., Fetter, R., Cardona, A., and Tomancak, P. (2012). Elastic volume reconstruction from series of ultra-thin microscopy sections. *Nature Methods*, 9:717–20.

Sánchez-Soriano, N., Bottenberg, W., Fiala, A., Haessler, U., Kerassoviti, A., Knust, E., Löhr, R., and Prokop, A. (2005). Are dendrites in *Drosophila* homologous to vertebrate dendrites? *Developmental biology*, 288(1):126–138.

Sato, T. (1968). A modified method for lead staining of thin sections. *J Electron Microsc (Tokyo)*, 17:158—-9.

Schaefer, J. E., Worrell, J. W., and Levine, R. B. (2010). Role of Intrinsic Properties in Drosophila Motoneuron Recruitment During Fictive Crawling. *Nature neuroscience*, 104(3):1257–1266.

Scott, E. K., Reuter, J. E., and Luo, L. (2003). Small gtpase cdc42 is required for multiple aspects of dendritic morphogenesis. *The Journal of neuroscience*, 23(8):3118–3123.

Seung, H. S. and Sümbül, U. (2015). Neuronal Cell Types and Connectivity: Lessons from the Retina. *Neuron*, 83(6):1262–1272.

Song, W., Onishi, M., Jan, L. Y., and Jan, Y. N. (2007). Peripheral multidendritic sensory neurons are necessary for rhythmic locomotion behavior in *Drosophila* larvae. *Proceedings of the National Academy of Sciences*, 104(12):5199–5204.

Suloway, C., Pulokas, J., Fellmann, D., Cheng, A., Guerra, F., Quispe, Stagg, S., Potter, C., and Carragher, B. (2005). Automated molecular microscopy: the new Leginon system. *Journal of Structural Biology*, 151:41–60.

Suster, M. and Bate, M. (2002). Embryonic assembly of a central pattern generator without sensory input. *Nature*, 416(6877):174–8.

Takemura, S.-y., Bharioke, A., Lu, Z., Nern, A., Vitaladevuni, S., Rivlin, P. K., Katz, W. T., Olbris, D. J., Plaza, S. M., Winston, P., et al. (2013). A visual motion detection circuit suggested by *Drosophila* connectomics. *Nature*, 500(7461):175–181.

Vazquez-Reina, A., Huang, D., Gelbart, M., Lichtman, J., Miller, E., and Pfister, H. (2011). Segmentation Fusion for Connectomics. *ICCV*.

Veeraraghavan, A., Genkin, A., Vitaladevuni, S., Scheffer, L., Xu, S., Hess, H., Fetter, R., Cantoni, M., Knott, G., and Chklovskii, D. (2010). Increasing Depth Resolution of Electron Microscopy of Neural Circuits Using Sparse Tomographic Reconstruction. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*.

Wilson, R. and Laurent, G. (2005). Role of GABAergic inhibition in shaping odor-evoked spatiotemporal patterns in *Drosophila* antennal lobe. *J Neurosci*, 25:9069–79.

Wu, Z., Sweeney, L., Chak, J. A. K., Andreone, B., Ohyama, T., Kerr, R., Luo, L., Zlatić, M., and Kolodkin, A. (2011). A combinatorial semaphorin code instructs the initial steps of sensory circuit assembly in the *Drosophila* CNS. *Neuron*, 70(2):281–98.

Zlatić, M., Landgraf, M., and Bate, M. (2003). Genetic specification of axonal arbors: *atonal* regulates *robo3* to position terminal branches in the *Drosophila* nervous system. *Neuron*, 37:41–51.

Zlatić, M., Li, F., Strigini, M., Grueber, W., and Bate, M. (2009). Positional cues in the *Drosophila* nerve cord: semaphorins pattern the dorso-ventral axis. *PLoS Biology*, 7(6):e1000135.

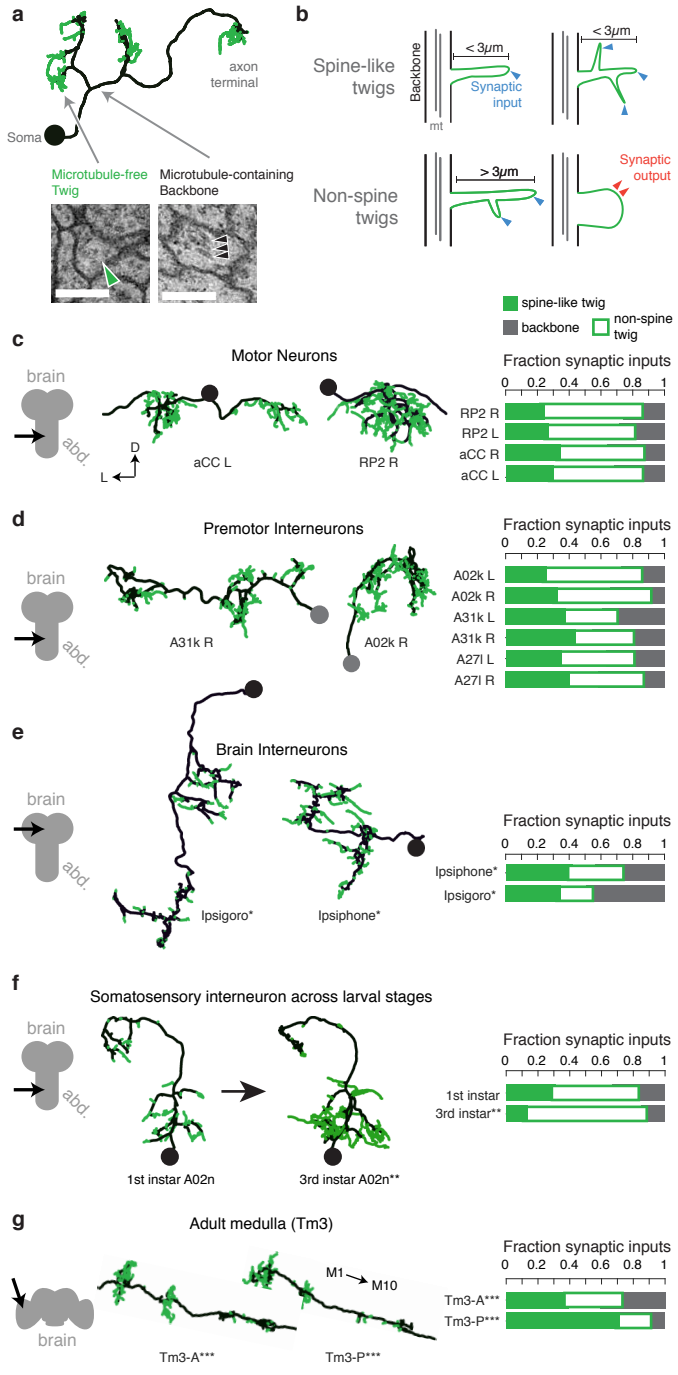

Figure 1: **Twigs, small microtubule-free neurites, are the primary site of input in *Drosophila* neurons.** **a)** A typical Drosophila neuron has a contiguous microtubule backbone from the soma to its arbors. Numerous terminal arbors without microtubules ("twigs") branch off the backbone. The presence (black arrows) or absence (green arrows) of microtubules can be seen in EM cross-sections of the neuron. **b)** Twigs less than 3 μm are considered spine-like, while those longer or primarily presynaptic are not. **(c–g)** EM reconstructions (middle) of *Drosophila* neurons from different parts of the nervous system (left) showing backbone (black) and twigs (green). At right, the fraction of all synaptic inputs onto short spine-like twigs, longer twigs, and backbone. Data sets are indicated by marks: no asterisk: 1.5 segment volume. *: Whole CNS volume. **: 3rd instar abdominal segment volume. ***: Adult medulla skeletons and images, generously provided by Janelia FlyEM [9]. Neurons are individually scaled to show complete arbors. **(c)** Motoneurons in 1st instar larva. **(d)** Premotor interneurons of 1st instar larva. **(e)** Interneurons in the brain of the 1st instar larva. **(f)** A somatosensory interneuron cell type across life stages, 1st instar and 3rd instar larvae. **(g)** Tm3 cells in the adult fly medulla.

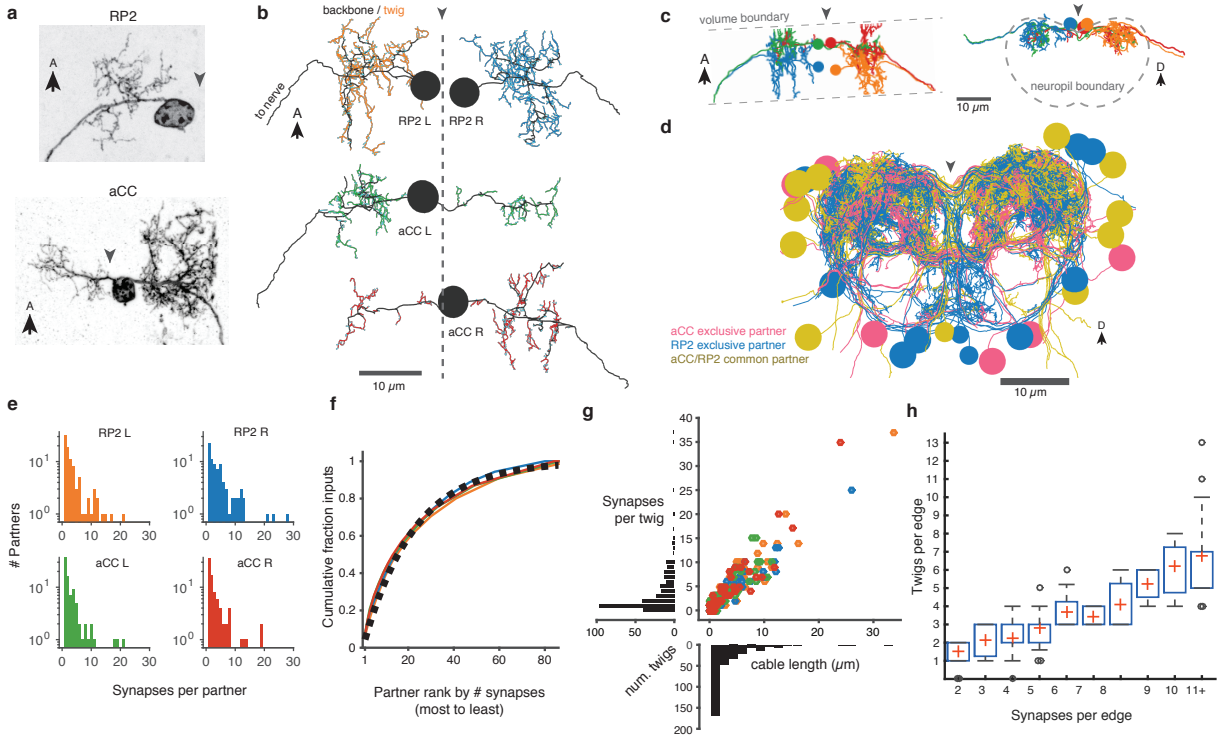

Figure 2: **Twigs are crucial to larval motor circuitry**. **(a)** Dorsal projections of genetically labeled motoneurons RP2 (top, from 1st instar) and aCC (bottom, from 3rd instar). Each cell type has characteristic dendritic arbors. Midline indicated by gray arrowhead. **(b)** EM reconstructions of each of four motoneurons aCC and RP2 in the 1st instar larva match the left and right homologs of aCC and RP2. Backbone is indicated by black, twigs by colors. Midline is shown as dashed line.**(c)** True spatial relationship of the four motoneurons in b), shown dorsally (left) and in cross-section (right). Note that the boundary of the EM volume is limited. **(d)** All arbors presynaptic to aCC and RP2. Colors indicate if neuron is presynaptic to one or both motoneuron cell types. **e)** Histograms of premotor partners connected via number of synapses. **(f)** Colored lines: the cumulative fraction of total inputs as a function of ranked presynaptic partners for each motoneuron are extremely similar. Black dashed line: simultaneous fit for all four motoneurons to $1 - \exp(-r/\rho)$ for rank r gives $\rho = 22.34$. **(g)** Scatterplot and histogram of the total length and number of synapses on each of the 305 twigs for each of the four motoneurons (colors as previous). **h)** Number of twigs contacted by motoneuron partners as a function of the number of synapses in the connection. Crosses are median, boxes the interquartile range, whiskers the 10th to 90th percentiles. Outliers shown.

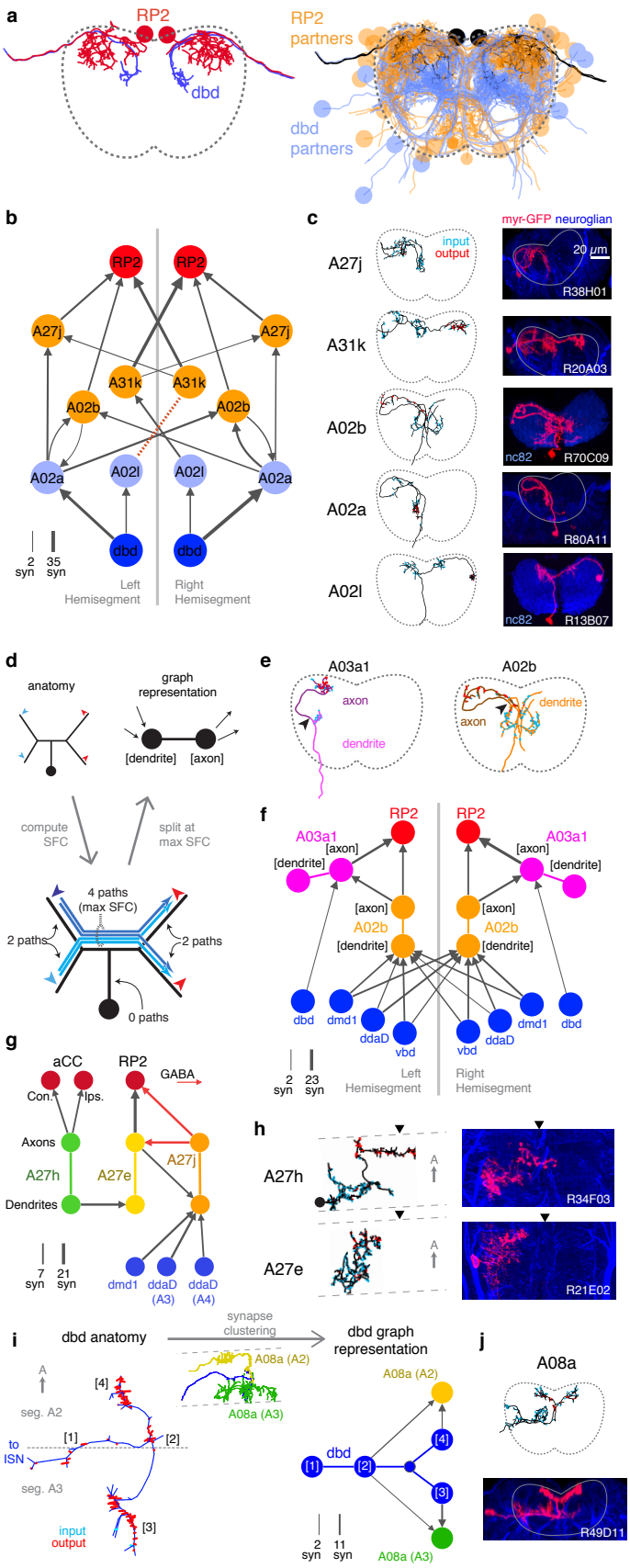

Figure 3: **Wiring diagrams for larval motor circuitry enriched with anatomical detail.** **a)** Starting with motoneuron RP2 and proprioceptive sensory neuron dbd (left), we identified all synaptic partners of each (right). **b)** Five symmetric pairs of identified neurons link the two cell types with three or fewer hops. All edges are observed in both the left and right hemisegments, except for a single edge outisde the volume boundary (red dashed line, see Supplemental Fig. 2). Line thickness increases with number of synapses (max. and min. values shown). In this and all network diagrams, single synapse edges are not shown for clarity. **c)** All identified cells in EM (left) could be matched to confocal maximum intensity projections of single neurons found in sparsely labeled GAL4 lines (right, see methods for details). For neuroglian staining, an approximate neuropile boundary is shown; for nc82 staining, the blue region is a profile of neuropile. **d)** Cartoon example of splitting neurons using synapse flow centrality (SFC). **e)** Examples of two premotor interneurons split into axonal (darker) and dendritic (lighter) regions with this method. Split point is indicated by the arrowhead. **f)** Splitting interneurons into axonal and dendritic compartments in a proprio-motor circuit reveals stereotypic pre- and post-synaptic connectivity to premotor interneuron A03a1 and differential contributions from proprioceptor dbd relative to other proprioceptors dmd1, ddaD, and vbd. **g)** Splitting interneurons A27j, A27e, and A27h reveals GABAergic pre- and post-synaptic input to a premotor connection, as well as dendro-dendritic coupling between interneurons that connect to synergistic motoneurons aCC and RP2. **h)** Dorsal projections of A27h and A27e from EM (left) and light (right), as in **c**. Midline indicated by arrowheads. **i)** Synapse clustering can represent dense groups of synapses on an arbor as distinct nodes in a graph widget. Here, each of four groups of synapses (numbers in brackets) is represented as a separate node in the graph. The graph demonstrates that posterior and anterior clusters (3 and 4, respectively) target homologous neurons in consecutive segments. **j)** Cross-sectional projection of A08a from EM (top) and light (bottom), as in **c**.

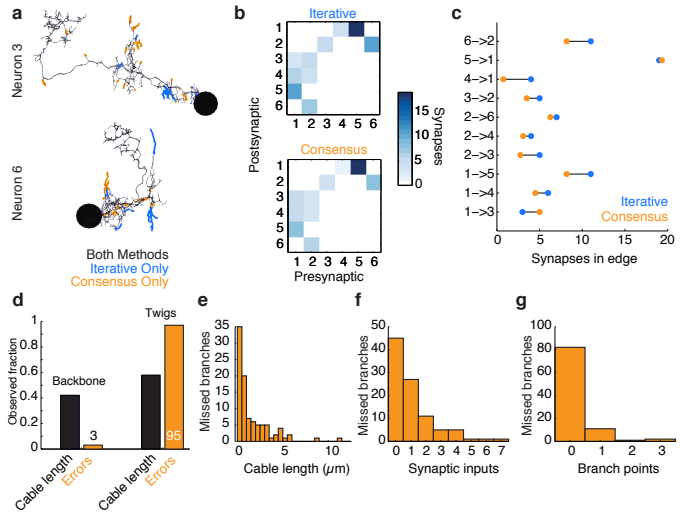

Figure 4: **Our method for iterative reconstruction produces reliable connectivity and anatomy. (a)** Dorsal view of two of six neurons for which we compared our iterative reconstruction method to a RESCOP-generated consensus of four independent reconstructions. Arbor found in both, dark lines; iterative only, blue; consensus only, orange. **(b)** The adjacency matrix produced by our iterative method has an identical set of edges as that of the consensus method, with variability only in the amount of synapses per edge. **(c)** The weights of each edge (the amount of synapses) are similar between methods. **(d)** Point errors in iterative reconstructions are not distributed equally across the cable of neuronal arbors, instead falling overwhelmingly on twigs. **(e-g)** Branches missed by our iterative method but observed in the consensus method are typically very small and lightly connected as seen from histograms of their **(e)** cable length, **(f)** synaptic inputs, and **(g)** number of branch points.

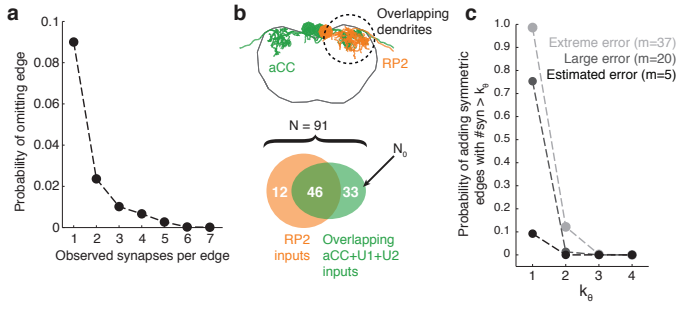

Figure 5: **Estimating errors that affect graph topology. a)** Estimated probability of fully omitting an edge as a function of how many synapses were on the edge based on omitting random twigs with the frequency observed in the validation data. **b)** Cartoon of dendritic overlap between RP2 and aCC, U1, and U2. On average, 91 axons put at least 2 synapses on any motoneuron (denoted N in the false positive estimate model, see text for details), of which 33 are not connected to RP2 (denoted N_0). **c)** Probability that, given a pair of homologous postsynaptic neurons, introducing m false inputs randomly distributed across N presynaptic neurons yields at least one pair of false edges of k_θ or more synapses each. The number of axons were estimated in **b**, and false input counts are shown estimated from the validation data ($m = 5$), as well as if they came from adding a rare but large twig ($m = 20$), and the largest observed twig ($m = 37$).

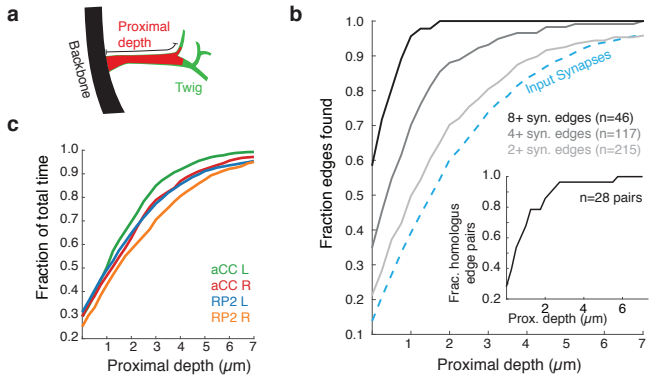

Figure 6: **Proximal regions of twigs reflect final wiring a)** Cartoon of the proximal depth (red) into of a twig (green) measured from from the backbone (black). **b)** The fraction of > 2 synapse edges onto aCC and RP2 that would be found when considering only synapses onto the backbone and twigs cropped at a maximum depth. From light to dark gray are those edges whose final measured connectivity has more than 2, 4, and 8 synapses. Blue dashed line indicates fraction of all input synapses. The inset depicts the fraction of pairs of homologous edges from identified neurons (N=28 edge pairs) that would be identified using synapses up to a given depth. **c)** Fraction of total reconstruction time for each of the four motoneurons (see legend) as a function of cropping twigs at a maximum depth. Note that 0 µm depth cropping corresponds to backbone reconstruction only.